

University of Groningen

Applied shotgun metagenomics approach for the genetic characterization of dengue viruses

Lizarazo, Erley; Couto, Natacha; Vincenti-Gonzalez, Maria; Raangs, Erwin C.; Velasco, Zoraida; Bethencourt, Sarah; Jaenisch, Thomas; Friedrich, Alexander W.; Tami, Adriana; Rossen, John W.

Published in:
Journal of Biotechnology: X

DOI:
[10.1016/j.btecx.2019.100009](https://doi.org/10.1016/j.btecx.2019.100009)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Lizarazo, E., Couto, N., Vincenti-Gonzalez, M., Raangs, E. C., Velasco, Z., Bethencourt, S., Jaenisch, T., Friedrich, A. W., Tami, A., & Rossen, J. W. (2019). Applied shotgun metagenomics approach for the genetic characterization of dengue viruses. *Journal of Biotechnology: X*, 2, [100009].
<https://doi.org/10.1016/j.btecx.2019.100009>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Applied shotgun metagenomics approach for the genetic characterization of dengue viruses

Erley Lizarazo^{a,1}, Natacha Couto^{a,1}, Maria Vincenti-Gonzalez^a, Erwin C. Raangs^a, Zoraida Velasco^b, Sarah Bethencourt^c, Thomas Jaenisch^d, Alexander W. Friedrich^a, Adriana Tami^{a,e}, John W. Rossen^{a,*}

^a University of Groningen, University Medical Center Groningen, Department of Medical Microbiology and Infection Prevention, Groningen, the Netherlands

^b Universidad de Carabobo, Facultad Experimental de Ciencias y Tecnología, Departamento de Biología, Valencia, Venezuela

^c Universidad de Carabobo, Facultad de Ciencias de la Salud, Departamento de Ciencias Fisiológicas, Unidad de Investigación en Inmunología, Valencia, Venezuela

^d University of Heidelberg, Heidelberg University Hospital, Department of Infectious Diseases, Section of Clinical Tropical Medicine, Heidelberg, Germany

^e Universidad de Carabobo, Facultad de Ciencias de la Salud, Departamento de Parasitología, Valencia, Venezuela

ARTICLE INFO

Keywords:

Shotgun metagenomics
Next-generation sequencing
Arboviruses
Dengue
Molecular epidemiology

ABSTRACT

Dengue virus (DENV), an arthropod-borne virus, has rapidly spread in recent years. DENV diagnosis is performed through virus serology, isolation or molecular detection, while genotyping is usually done through Sanger sequencing of the envelope gene. This study aimed to optimize the use of shotgun metagenomics and subsequent bioinformatics analysis to detect and type DENV directly from clinical samples without targeted amplification. Additionally, presence of DENV *quasispecies* (intra-host variation) was revealed by detecting single nucleotide variants. Viral RNA was isolated with or without DNase-I treatment from 17 DENV (1–4) positive blood samples. cDNA libraries were generated using either a combination of the NEBNext[®] RNA to synthesize cDNA followed by Nextera XT DNA library preparation, or the TruSeq RNA V2 (TS) library preparation kit. Libraries were sequenced using both the MiSeq and NextSeq. Bioinformatic analysis showed complete ORFs for all samples by all approaches, but longer contigs and higher sequencing depths were obtained with the TS kit. No differences were observed between MiSeq and NextSeq sequencing. Detection of multiple DENV serotypes in a single sample was feasible. Finally, results were obtained within three days with associated reagents costs between €130–170/sample. Therefore, shotgun metagenomics is suitable for identification and typing of DENV in a clinical setting.

Introduction

Dengue viruses (DENV) belong to the *Flaviviridae* family and are among the most widely distributed arthropod-borne viruses worldwide. All DENV cause dengue fever, a self-limited febrile illness but in some patients, dengue becomes a life-threatening illness [1]. In the last five decades DENV has rapidly spread around the globe. This together with high morbidity rates make DENV a public health threat in tropical and subtropical regions and increasingly in temperate countries [2–4].

DENV are single-stranded, positive sense RNA viruses with a genome of approximately 10.7 kb that contain a single open reading frame (ORF) [5]. They comprise 4 antigenically distinct serotypes (DENV-1 to 4) that have up to 65% genome sequence identity [6] and cluster into different genotypes as a result of high mutation rates in

their genomes [2,7]. Disease outcome and virus transmission rates have shown to be genotype-dependent [8].

Diagnosis of DENV can be performed by serological testing, isolation of the virus or through molecular methods [9]. Genotyping is often based on Sanger sequencing (parts) of genes encoding the structural proteins, mostly the envelope (E) gene [10] or, alternatively the Capsid pre-membrane CprM gene [11,12]. DENV genotypes are defined as clusters with associations on epidemiological grounds with a sequence divergence of $\leq 6\%$ [13,14]. Using Sanger sequencing to sequence the whole genome requires amplification of multiple overlapping fragments [15–19], it is time consuming and not suitable for high-throughput. Nevertheless, it reveals phylogenetic relationships at the highest resolution, enables detection of recombinant events and escape mutants, and results in a better understanding of the dynamics of DENV

* Corresponding author at: Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, Hanzeplein 1 (HPC EB80), 9713 GZ Groningen, the Netherlands.

E-mail address: j.w.a.rossen@rug.nl (J.W. Rossen).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.btecx.2019.100009>

Received 12 January 2019; Received in revised form 5 May 2019; Accepted 6 May 2019

Available online 15 May 2019

2590-1559/© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

evolution and its implications in disease development [18]. Moreover, it has been proposed that genetic variants of RNA viruses, named “*quasispecies*”, present in the host may influence pathogenesis and disease outcome of RNA viruses in human infections [20–22]. A more rapid and cost-effective way to obtain these data may be the use of shotgun metagenomics that can be applied for all viruses in all kinds of clinical material. In clinical microbiology laboratories, short-read sequencing (SRS) is still the most frequently used method, although long-read sequencing slowly finds its way. For SRS, cDNA first needs to be fragmented for which mainly two methods are available, i.e., enzyme-based fragmentation or mechanical shearing. The often used Nextera XT DNA [NXT] library preparation kit, e.g., fragments cDNA using a patented transposon/transposase-mediated cleavage mechanism, with DNA fragments subsequently being amplified using primers targeted to adaptor sequences linked to the transposon [23].

In contrast, for the TruSeq (TS) V2 DNA or RNA kit, e.g., [TS] the cDNA is first fragmented by mechanical shearing, followed by end-repair of the fragments and adaptor ligation [24]. The NXT kit has the advantage that it requires only 1 ng of input cDNA and has a significantly faster preparation time after cDNA synthesis compared to the TS method [24]. However, GC bias can have a prominent impact on transposase-based protocols, like the NXT, likely through a combination of transposase insertion bias being coupled with a high number of PCR enrichment cycles. Apart from the library preparation method, different sequencers may have different sequencing errors that might influence the final sequence quality [25]. In this study, we evaluated the use of shotgun metagenomics and bioinformatics analyses to detect and type DENV and to reveal the presence of DENV *quasispecies* directly from sera and plasma samples. In addition, we assessed the effect of: i) DNase I treatment to decrease the human DNA background (to increase the number of reads belonging to viruses and potentially the sensitivity of the approach); ii) two different library preparation methods and iii) two sequencing platforms, on the sequence data quality. For this we a) performed identification and molecular characterization of DENV in the tested samples, b) performed phylogenetic analysis using nearly entire genomes, and c) screened for DENV *quasispecies* (intra-host variation) by detecting single nucleotide variants (SNVs).

Materials and methods

Sample collection

Plasma (n = 9) or serum samples (n = 8) from seventeen confirmed dengue symptomatic patients were collected in Venezuela between 2010–2015 (Table 1). DENV positivity was confirmed by either RT-qPCR (CDC, 2013) or nested RT-PCR [9].

RNA isolation and DNase I treatment

Viral RNA was isolated from 140 µL of serum or plasma using the QIAamp viral RNA isolation kit (Qiagen, Hilden, Germany). To reveal if DNase I was able to decrease the human DNA background [26] and thereby increase the sensitivity to detect DENV, 12 out of 17 samples were divided into 2 aliquots. One aliquot of the sample was treated with RNase-Free DNase I (Qiagen) for 15 min using an the optional on-column DNA digestion during the QIAamp isolation, while the other followed the regular QIAamp viral RNA isolation kit protocol. The RNA of the remaining 5 samples was extracted with the on-column DNA digestion step during the isolation procedure.

To control for possible contaminations, negative controls (DNA- and RNA-free water, Sigma-Aldrich, St. Louis, MO, USA) were included. As a positive control, we used the supernatant of a viral culture containing DENV-2 strain 16,681. In addition, to test if multiple DENV serotypes could be detected simultaneously, a single sample was spiked with a combination of positive patient specimens infected with DENV-1, -2 and -3 prior to library preparation.

Library preparation

Two commercial kits were used for library preparation prior to sequencing: a) the TruSeq RNA V2 (TS) and b) Nextera XT DNA (NXT) kits, both from Illumina (San Diego, CA, USA). For the TS preparation protocol, the poly-A purification step was omitted as the poly-A tail is lacking in DENV [27]. Thus, we started by adding 5 µL of eluted RNA (5–10 ng/µL) directly to the fragmentation step and then followed the instructions of the manufacturer. For NXT, prior to the library preparation, the eluted RNAs were cleaned with the Agencourt RNAClean XP (Beckman Coulter, Brea, CA, USA) system. Next, cDNA was synthesized using the NEBNext[®] RNA First and Second strand modules (New England Biolabs, Ipswich MA, EUA). The cDNA was purified using the QIAquick PCR Purification Kit (Qiagen). Subsequently, 1 ng of cDNA was used for the NXT DNA library preparation and the subsequent steps followed the manufacturer's protocol.

For both TS and NXT methods, quality controls were performed during and after library preparation using the Qubit 2.0 fluorometer (Life Technologies, Thermo Fisher Scientific, Carlsbad, CA, USA) and by a 2200 TapeStation (Agilent Technologies, Waldbronn, Germany). A size selection (1:1 beads/sample ratio) of the libraries with AMPure Beads (Beckman Coulter, Brea, CA, USA) was conducted to discard unwanted adapter dimers in order to ensure optimal results.

Next-generation sequencing (NGS)

NGS was performed by combining 12 or 24 libraries in equimolar ratios before loading them on a MiSeq or on a NextSeq 500 sequencer (Illumina, San Diego, CA, USA), and at 12 pM and 1.8 pM, respectively. The MiSeq Reagent Kit V2 and the NextSeq Series Mid-Output kit (Illumina) were used to generate 150-bp paired-end reads. MiSeq data were processed with MiSeq control software v2.4.0.4 and MiSeq Reporter v2.4 and NextSeq data with bcl2fastq2 conversion software v2.18 (Illumina).

Data analysis

To identify, genotype and characterize DENV in the tested samples, the fastq files were analyzed employing two different approaches. First, paired-end reads were uploaded to Taxonomer (IDbyDNA, San Francisco, CA, USA), a web-based metagenomics freely available analysis tool. In short, reads were analyzed through the integrated tools: Binner, Classifier, Protonomer and Afterburner to identify micro-organism communities and results were visualized through <http://taxonomer.io.bio.io> [28].

The second approach, used the CLC Genomics Workbench v10.1.1 software (Qiagen). The workflow used (See Fig. S1 and Table S1) started with quality assessment of the reads and subsequent quality trimming of unwanted adapters using a limit of 0.05 prior to mapping the reads against the human genome (hg18). Then the unmapped reads were collected for *de novo* assembly. The consensus sequence of the longest contig was extracted and used for viral identification using blastn. Additionally, to facilitate the generation of whole genomes, the unmapped reads were also used to map against prototype DENV strains retrieved from GenBank (See Table S2). To detect the ORF, we utilized the CLC Genomics Workbench plugin MetaGeneMark v1.4 (Gene Probe). In addition, the presence of DENV *quasispecies* was examined by analyzing presence of SNVs by the Low Frequency Variant Detection module from CLC Genomics Workbench v10.1.1 which includes i) a statistical model for SNV calling that relies on a multinomial analysis to determine the presence of different variants at a given site of the analyzed sample and ii) an error model to account for sequencing errors. In the SNV calling workflow we have used for each sample the consensus sequence of the DENV found in it as reference. The selected cut-offs for SNV calling were 500-fold coverage, 1% SNV frequency, and the required significance started at 5% (for detailed information about the

Table 1
Description of Venezuelan samples sequenced in this study.

Sample ID	Sample type	Origin	Age of patient (years)	Collection Date	Serotype	Days after symptoms onset	Type of infection ^y	Viral RNA (copies/ μ L)	Clinical classification [§]
ID01	Serum	Aragua	17	27/08/2010	DENV-3	3	*	*	a
ID02	Serum	Aragua	7	30/08/2010	DENV-1	2	*	*	b
ID03	Serum	Aragua	18	31/08/2010	DENV-2	3	*	*	b
ID04	Serum	Aragua	13	01/09/2010	DENV-4	3	*	*	a
ID05	Serum	Aragua	21	07/09/2010	DENV-1	3	*	*	b
ID06	Serum	Aragua	8	16/02/2011	DENV-4	3	*	*	a
ID07	Serum	Aragua	50	01/11/2011	DENV-4	3	Secondary	*	b
ID08	Serum	Aragua	21	17/07/2012	DENV-4	3	*	*	b
ID09	Plasma	Carabobo	11	22/09/2015	DENV-2	3	Probable secondary	1,430	b
ID10	Plasma	Carabobo	9	23/09/2015	DENV-3	2	Probable secondary	5	b
ID11	Plasma	Carabobo	17	25/09/2015	DENV-1	4	Probable secondary	82,600	b
ID12	Plasma	Carabobo	6	30/09/2015	DENV-3	3	Inconclusive	603	b
ID13	Plasma	Carabobo	18	05/10/2015	DENV-3	2	Probable secondary	1,070,000	b
ID14	Plasma	Carabobo	18	06/10/2015	DENV-1	2	Probable secondary	44,300	b
ID15	Plasma	Carabobo	17	15/10/2015	DENV-2	3	Probable secondary	2,870	b
ID16	Plasma	Carabobo	33	19/10/2015	DENV-1	3	Probable secondary	191	b
ID17	Plasma	Carabobo	14	27/10/2015	DENV-2	2	Probable secondary	6,600	b

Abbreviations: DENV, dengue virus.

^yClassification of DENV type of infection according to Cordeiro et al. [55].

[§]Data not available.

^aClassification of dengue with/without warning signs according to WHO [56] (available at <http://www.who.int/rpc/guidelines/9789241547871/en/>).

^bDengue with warning signs.

^cDengue without warning signs.

workflow parameter see Table S1).

To assess if multiple detection of DENV serotypes in a single sample was feasible we performed an *in-silico* validation analysis to evaluate the ability of CLC Genomics Workbench to discriminate between DENV in mixed samples. For this, fastq files containing the raw reads of each of the four DENV serotypes (samples: ID06, ID09, ID13, ID14; Table 1) were merged into a single file and assessed through our CLC genomics workbench pipeline (Fig. S1). In addition, we mixed RNA from DENV 1, 2 and 3 isolated from positive tested samples before library preparation and sequencing. Subsequent, data analysis was performed through the aforementioned workflow. In both approaches, the web-based tool Taxonomer (IDbyDNA, San Francisco, CA, USA) was applied for the fast identification of DENV from the raw reads.

To determine the phylogenetic relationship of the newly generated genomes, the obtained sequences were aligned against genomes of known genotypes retrieved from GenBank (See Table S3). The multiple nucleotide sequence alignment was performed with MAFFT v7.313 [29]. The sequence alignment was edited manually to generate an alignment with ORF only. A maximum-likelihood (ML) phylogenetic tree was estimated using RAxML [30] under general time reversible model with gamma-distributed rates distribution substitution model (GTR+ Γ), which was determined as the best-fit model using CLC Genomics Workbench (data not shown).

Statistical analyses were performed using SPSS v23 (IBM, New York, United States of America). The Wilcoxon signed-rank non-parametric test was used to detect significant differences between continuous (i.e. number of reads) and categorical variables (i.e. library preparation kits). Significance was determined at the 5% level (p -value ≤ 0.05).

Results

Effect of DNase I treatment

From twelve out of seventeen samples, two aliquots were obtained and extracted with and without DNase I treatment to reveal its effect on the number of human and DENV reads. The average total number of reads without DNase I treatment per sample for NXT was 2,090,050 of which 938,543 (45%) mapped against the human genome, which is lower than the average of 6,201,468 total reads obtained with TS of which 5,484,689 reads (80%) mapped against the human genome (Fig. 1). The number of human reads decreased significantly ($p < 0.05$), on average by 18% for both NXT and TS, using the DNase I treatment. The depletion of human DNA had a positive effect in the proportion of reads that matched DENV genomes (Fig. 2). After the treatment with DNase I, an average increase in the number of reads matching DENV of 313 and 39 times was observed for the NXT and TS

approach, respectively. Few DENV reads were identified in the negative control (0.004%–0.006% of the reads). The DENV-2 strain 16,681 was successfully identified in the positive control.

Next, we compared the two different library preparation methods (NXT and TS) in combination with two different NGS platforms (MiSeq and NextSeq). A summary of the quality parameters for the different runs is shown in Table 2. Optimal raw cluster density for MiSeq using a v2 cartridge has been reported to be between 1,000–1,200 K/mm² and for NextSeq between 170–220 K/mm² [31]. In our study, two runs had raw cluster densities under the desirable range (869 K/mm² for MiSeq, 22 ± 4 K/mm² for NextSeq), however, the quality scores (Q30) of both runs were higher than those runs with optimal raw cluster densities (Table 2).

Table 3 shows a summary of the results obtained for DENV with the different sequencing approaches. We compared the average depth coverage and the average contig length after *de novo* assembly and after mapping. The libraries prepared with TS showed a higher average depth coverage. Yet, all runs resulted in an average depth coverage of $> 1,800$ -fold. Likewise, both library preparation methods resulted in complete or nearly complete DENV genomes when reads were mapped against the reference genomes. However, TS provided slightly longer contigs when performing *de novo* assembly compared to NXT. The DENV serotypes identified through CLC Genomics Workbench from the reads obtained by NXT/TS and MiSeq/NextSeq were 100% concordant with the results of the RT-PCR or RT-qPCR. However, when Taxonomer was used to analyze samples prepared with NXT and ran on NextSeq, the DENV serotypes of two samples did not match those identified by RT-PCR and CLC Genomics Workbench v10.1.1.

Additionally, we examined the genome wide depth coverage and the G/C content of the sequenced genomes with NXT and TS (an example is shown in Fig. S2), in an attempt to identify areas of low coverage due to variable G/C content. The results showed a good proportion of reads per base position and a comparable G/C pattern in both preparations. Nevertheless, the 5' and 3' regions had the lowest coverage and were the most variable sections to be sequenced and, consequently, caused differences in the length of the genomes. The consensus of assembled/mapped genomes were a few nucleotides (nt) shorter than the references in the GeneBank (Table S2). In the case of NXT an average of 24 ± 37 nt and 40 ± 32 nt in the 5' and 3' ends, respectively, were missing. Whereas for the TS, on average 14 ± 30 nt and 23 ± 27 nt were missing in the 5' and 3' ends, respectively. When comparing the time investment, both methods performed similarly. Thus, the whole workflow from the viral RNA isolation to genome assembly through both library preparation methods and sequencing platforms took approximately three days.

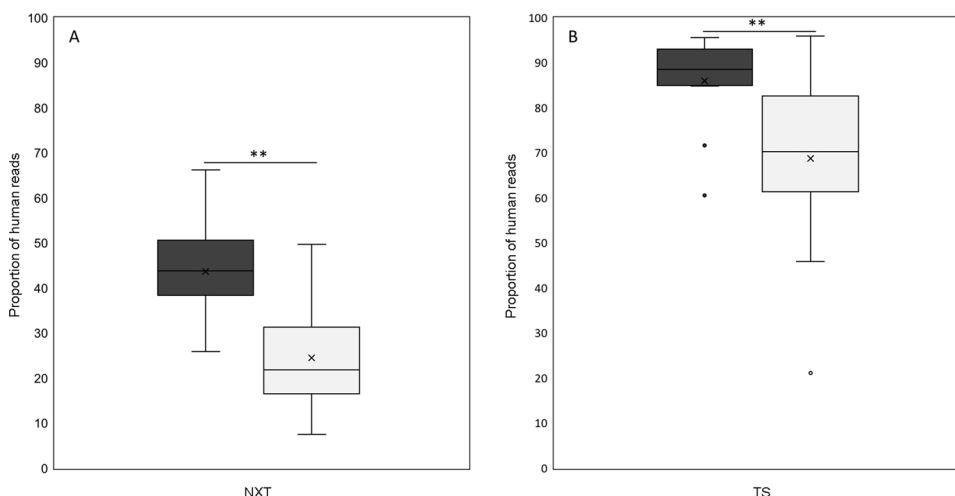


Fig. 1. Boxplot showing the effect of DNase I (Qiagen) treatment in the proportion of human reads. Using the NXT library preparation kit (A) or the TS kit (B). The proportion of human reads was calculated as the number of human reads over the total number of reads for each sample. Grey bars show the DNase I treated samples whereas black bars show the DNase I untreated samples. Upper and bottom whisker lines represent 1.5 interquartile range (IQR); the box represents the upper and lower quartiles, horizontal line within the box represent the median and the x denotes the arithmetic means; dots denotes outliers. **, p -value < 0.05 (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

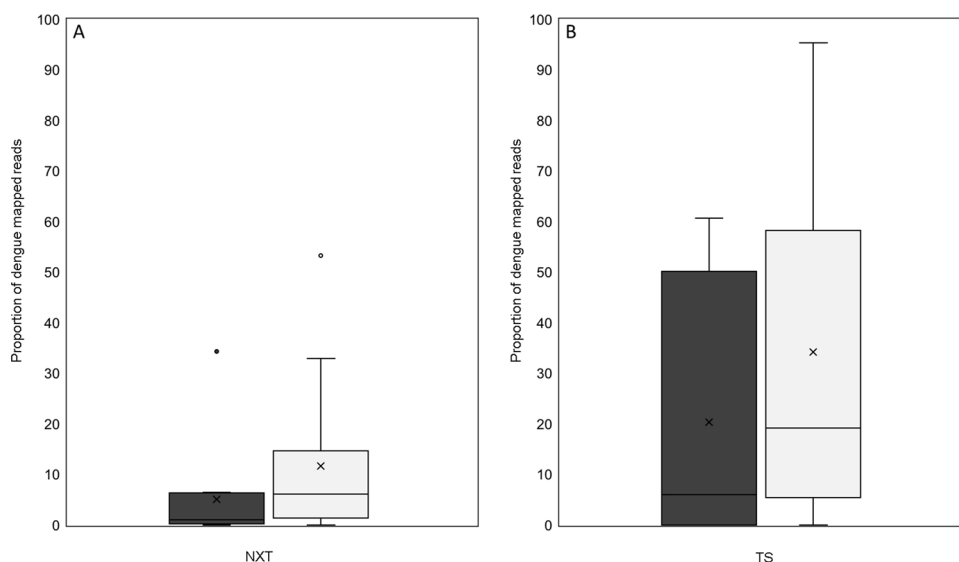


Fig. 2. Boxplot showing the effect of DNase I (Qiagen) treatment on the proportion of mapped DENV reads. Using the NXT library preparation kit (A) or the TS kit (B). The proportion of DENV reads was calculated as the number of reads that mapped DENV over the total reads of each sample. Grey bars show the DNase I treated samples whereas black bars show the DNase I untreated samples. Upper and bottom whisker lines represent 1.5 inter-quartile range (IQR); the box represents the upper and lower quartiles, horizontal line within the box represent the median and the x denotes the arithmetic means; dots denotes outliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 2

Sequence quality of the 4 runs performed using two different library preparation kits and two sequencing platforms.

Platform	Library Prep	Raw density (K/mm2)	%PF	% ≥ Q30	Total reads	Total reads (PF)	Yield Gbp
MiSeq	NXT	1,082 ± 38	86.12	82.96	40,332,330	34,734,340	5.45
MiSeq	TS	869 ± 16*	91.51	92.95	32,919,292	30,123,438	4.59
NextSeq	NXT	22 ± 4*	98.82	96.50	38,684,052	37,613,244	2.29
NextSeq	TS	179 ± 4	89.75	84.08	114,917,472	103,134,945	41.66

Abbreviations: Gbp, giga base pair; PF, passing filter; Q30, quality score with base call accuracy of 99.9% (1 incorrect base in 1000 based calls); NXT, Nextera XT library prep; TS, TruSeq v2 RNA library prep.

* Raw density was under the optimal range.

Detection of multiple DENV serotypes in a spiked sample

The workflow correctly identified the presence of multiple DENV serotypes in the spiked sample and in the *in silico* sample (Table 4 and 5, respectively). In the spiked sample, the proportions of DENV varied from as low as 0.02% for DENV-1 to as high as 3.12% for DENV-3. The genomes had a coverage depth of between 17- to 728-fold. However, complete genomes could not be *de novo* assembled for all DENV, instead a nearly complete genome (10,555 bp) was generated for DENV-3 and shorter assembled contigs were generated for DENV-1 and DENV-2 (2,796 bp and 6,736 bp, respectively; Table 4). Therefore, we mapped the reads against reference genomes (See Table S2) to generate longer consensus with nearly full genomes of the three DENV being generated (10,614 bp for DENV-1; 10,675 bp for DENV-2 and 10,675 bp for DENV-3). Similar results were obtained during the *in silico* detection. All DENV serotypes in the simulated specimen showed a correct identification through the CLC Workbench workflow. Likewise, the generation of nearly complete genomes for all four serotypes was achieved using the mapping approach

(10,700 bp for DENV-1; 10,691 bp for DENV-2; 10,673 bp for DENV-3 and 10,637 bp for DENV-4). On the other hand, as shown in Table 5, *de novo* assembly generated contigs for all DENV serotypes but failed to assemble the entire ORF of DENV-1 (5,391 bp). Nonetheless, despite the notable differences of reads' abundance per serotype in the simulated sample (ratios: DENV-4/DENV-3 [40:1]; DENV-4/DENV-2 [21.7:1]; and 57.7:1 for DENV-4/DENV-1) all DENV were detected and nearly complete genomes were obtained.

Detection of SNVs in DENV

Twelve out of 17 samples matched the criteria used for SNVs calling.

The non-synonymous SNVs identified in each sample are shown in Table S4 and the frequency and position of each change is shown in Fig. 3. For DENV-1, we identified non-synonymous SNVs in all samples tested. Around 75% of the SNVs of DENV-1 generated a frame shift at different positions of the polyprotein (average frequency of ~3% in the sequenced reads), the SNV with higher frequency (24.1%) was detected in sample ID11 in the E protein encoded region (Val708Met). Additionally, an aggregation of SNVs in the NS5 encoded region was detected. In DENV-2 the non-synonymous SNVs were found in two of the tested samples with an average of 58% of SNVs generating frame shifts (average frequency of ~4%). One of the SNVs was placed in the M encoded region, while the majority occurred in the NS encoded regions. Likewise, in sample ID09 one SNV generated an early stop codon on position 3,237 (Trp > Stop). In DENV-3 we detected six non-synonymous changes in one sample (ID01). However, these changes did not include frame shifts or stop codons. In DENV-4, the SNVs were detected in the capsid, envelope and NS encoded regions. On average, 59% of the SNVs detected caused a frame shift (average frequency of ~3%).

Phylogenetic characterization of DENV

Phylogenetic trees generated from the complete ORFs (Fig. 4) showed that the isolates of DENV-1 clustered within genotype V with some closely related isolates from Colombia 2008 (GQ868570) and Ecuador 2014 (MF797878). The four DENV-2 isolates fell within the American/Asian cluster genotype, a genotype often associated with disease severity [32]. All DENV-3 isolates clustered within genotype III, and were related mainly to other Venezuelan isolates. The DENV-4 strains fell within genotype II and clustered into two different groups. The isolated DENV strains were closely related to Venezuelan isolates, and for every serotype only one genotype was detected.

Table 3
Comparison of the effect of different library preparation methods and sequencing platforms in the proportion of DENV reads.

Platform	Library Preparation	No. Samples	Average total number of reads	Average human mapped reads	Average un mapped Reads	Average DENV mapped reads	Average proportion of DENV mapped*	Average depth coverage (fold)	Average assembled consensus length (bp)	Average mapped consensus length (bp)
MiSeq	NXT	5	2,792,515	2,042,971	749,544	469,950	63%	5,603	7,667	10,683
	TS	5	1,396,358	1,182,726	213,632	136,998	64%	1,867	10,202	10,680
NextSeq	NXT	12	677,692	155,878	521,814	175,820	34%	2,498	4960	9,543
	TS	12	6,179,416	4,054,813	2,124,603	1,547,456	73%	21,327	10,347	10,483

Abbreviations: bp, base pair; NXT, Nextera XT library prep; TS, TruSeq v2 RNA library prep.

* Using the number of unmapped reads as denominator.

Discussion

Metagenomics studies have proved their value in clinical diagnostic settings and for surveillance [33,34]. Here, we applied a high-throughput NGS assay for direct whole-genome sequencing of DENV directly from clinical samples. This method avoids the need to design primers to type, thus allowing for unbiased typing and, therefore, is able to identify uncommon variants or those that would be missed if a primer-based method is used. Therefore, it has more discriminatory power than methods that target specific regions. Moreover, co-infection with different DENV serotypes or with other microorganisms can be detected in a single reaction. The entire procedure described took approximately 3 days to complete. The library-associated cost was €57 per sample using NXT and €74 per sample using TS (date of cost assessment: January 2018). The run-associated cost was €86 per sample if 12 samples were multiplexed in one MiSeq run or €70 per sample if 24 samples were multiplexed in one NextSeq run (date of cost assessment: January 2018). However, these costs do not include the investment, service cost and personnel associated with each NGS sequencing platform. The overall costs are comparable to those of Sanger sequencing [35]. Nonetheless our method, contrary to the latter, allows for sample multiplexing and is able to detect low frequency variants and co-infections [34] making it more cost-effective in a diagnostic setting.

The sequencing output of the shotgun metagenomics approach depends among other factors, on the amount of viral RNA present in the sample and also on the amount of human DNA; the latter affecting the yield and the sensitivity of the protocol as it also serves as a template during library preparation. As a result, viral sequence depth could vary depending on the total nucleic acid yield [36]. Therefore, we tested the effect of DNase I treatment on sequencing outcomes [26]. The results of the DNase treatment showed to be different between NXT and TS, which can be attributed to the requirements of additional steps during cDNA synthesis (RNA bead cleaning and cDNA purification) prior to the NXT library preparation. However, we were unable to assess the residual DNA present after DNase treatment in order to estimate the efficiency of the DNase on each sample. Nonetheless, DNase I treatment appeared to be effective in decreasing the human DNA background and increasing the yield of DENV reads in both the NXT and TS approaches. Similar findings have also been described for polioviruses, where DNase-treatment significantly increased the percentage of reads mapped to the targeted poliovirus genomes compared with that from non-DNase treatment [37]. DNase treatment allows a higher number of samples to be multiplexed and sequenced in a single run with the required sequence depth (> 500-fold) thereby reducing the cost per sample.

Both NXT and TS can be used for DENV sequencing, nonetheless the average depth coverage along with the whole genome was higher and more homogeneous using TS compared to NXT. Similar results were described in a previous study, showing that the input DNA quality had no effect on TS data (i.e. depth coverage), but had a significant effect on NXT data [23], meaning that a DNA sample of lower quality had a worse impact on the NXT libraries than on the TS libraries. This might also explain why the assembled consensus obtained from NXT libraries were divided into small contigs, while the ones from TS were consistently longer. Yet, this limitation can be surmounted by performing mapping with reference strains instead of *de novo* assembly. Another problem, however, was that even after mapping, the contigs obtained were, on average, smaller using the NXT. This might be due to limitations of the library preparation during fragmentation, as the NXT kit uses tagmentation for this purpose while the TS kit uses mechanical fragmentation. However, this could also have been caused by the low raw cluster density in the NXT-NextSeq run and further studies are needed to confirm this observation.

The advantages of using whole-genome sequences compared to partial sequences for phylogenetic analysis have been shown previously and include correct identification of outbreak strains due to its

Table 4
Results obtained from a spiked sample with multiple DENV serotypes.

Virus	Total reads	<i>de novo</i> assembly				mapped assembly against DENV reference genomes			
		Mapped reads	Percentage mapped reads (%)	Depth coverage	Longest consensus (bp)	Mapped reads	Percentage mapped reads (%)	Depth coverage	Consensus (bp)
DENV-1	1,783,278	368	0.02%	18	2,796	1,180	0.07%	15	10,614
DENV-2	1,783,278	1528	0.09%	31	6,736	2,514	0.14%	32	10,675
DENV-3	1,783,278	55,661	3.12%	728	10,555	55,952	3.14%	734	10,675

Abbreviations: bp, base pair.

increased discriminatory capacity [38]. In this study, we were able to highly discriminate between strains that belonged to the same genotype. Although, for each DENV serotype only one genotype was detected, our isolates appear to cluster within distinct subpopulations, which could be related to the extensive DENV genetic variability or to multiple introductions of different subpopulations in the country as reported earlier [39,40]. For instance, our DENV-1 and DENV-3 isolates showed high identities with isolates from other Latin American countries. This may be explained by the movement/migration/travelling of people between these countries, for example from Colombia to Venezuela in the last 50 years [41].

The applied protocols showed a high sensitivity and specificity (up to 100%) when compared to RT-PCR or RT-qPCR, and were able to detect DENV in clinical samples with as low as 5 viral copies/μL. Taxonomer was used as a first approach for rapid detection of DENV (5–10 minutes) however, it failed in two samples reporting instead several serotypes, including the correct one but in a low proportion. This may be explained by the low amount of reads in these samples and by the nature of Taxonomers' kmer search (parameters: 6-frame translation; kmer size 30; 10 amino acids), whereby if reads that belong to a shared DENV genome region are mainly found, the chances of false positives are higher [28]. This, however, was overcome by using the CLC Genomics Workbench approach, which had 100% concordance with the PCR results. Likewise, as shown in the *in silico* assay the shotgun metagenomics workflow was able to detect multiple DENV in a single sample (spike-in sample) without targeting any specific serotype, which surmounts challenges like template concentration, sequence diversity, primer specificity, and PCR amplification efficiency. These challenges have been reported in previous attempts to sequence multiple DENV with targeted full-genome amplification and sequencing either by Sanger or amplification-based NGS approaches [16,17]. Likewise, the ability to detect multiple DENV serotypes together with the high throughput of the NGS platforms could facilitate the in-depth analysis of co-viral infections and their possible clinical manifestations.

Another advantage of NGS is the study of inter- and intra-host relations of viral genetic variants [34]. The advantage of this approach is that no specific amplification is required, which represents an unbiased approach to screen for natural mutations across the DENV genome within the host. We were able to detect SNVs in 71% of our samples. DENV-1 strains isolated in different time and geographical points had similar frame shifts and overall shared SNVs through their genomes. In

addition, in DENV-1 isolates more SNVs could be detected and were more frequent than in other DENV serotypes, suggesting a different stage of diversification. Some SNVs detected in DENV-1 and other serotypes represented multiple deleterious mutations such as frame shifts, intragenic stop-codons, nucleotide insertions or deletions that could affect viral pathogenesis by generating defective viral particles [42,43]. In concordance with our findings, deleterious mutations were reported to be transmitted together with wild-type viruses of DENV-1 in Myanmar [44]. Moreover, it was proposed that the defective genomes were acting as defective interfering viral particles that resulted in attenuation of disease severity, increasing the spread of the virus by allowing greater mobility of human hosts [45]. However, more studies are needed to confirm these observations in our population. Thus, epidemiological data linked to unbiased deep whole genome sequencing data can reveal a specific change in viral fitness or clinical disease development during DENV transmission, in a fraction of the time taken by other approaches [18].

One of the major limitations of this study was the different raw densities obtained from the four sequencing runs shown in Table 3, which were especially low for the NextSeq-NXT run. Although, the low densities of Next-NXT resulted in lower number of DENV reads, and consequently in lower depth coverage and shorter contigs, the run still produced enough reads to enable the fast detection of DENV through Taxonomer (with only two misclassifications) and the generation of nearly complete genomes through mapping to DENV references. In order to minimize the possibility of inconsistent cluster generation it is recommended to perform an extra step of size selection of small fragments (i.e. indexes, primers) after the pooling of libraries step. If this step is not performed, such small fragments can generate both background noise and losses in sequencing depth.

As with other (molecular) methods several controls should be included to validate the obtained results, including a negative control. In our negative control, we detected DENV reads, although it represented only 0.004%–0.006% of the reads. These results may be due to contamination during library preparation (e.g. sample-to-sample contamination prior to indexing), the result of sequencing artefacts (e.g. demultiplexing errors, sample bleeding), or to incorrect classification during data analysis (e.g. highly homologous regions) [46]. Our samples and sequencing libraries were handled in laminar flow cabinets; however, we cannot exclude the possibility of contamination. Furthermore, the reagents used may also be or become contaminated with

Table 5
Results obtained from the *in silico* analysis of the simulated specimen.

Virus	Total reads	<i>de novo</i> assembly				mapped assembly against DENV reference genomes			
		Mapped reads	Percentage mapped reads (%)	Depth coverage	Longest consensus (bp)	Mapped reads	Percentage mapped reads (%)	Depth coverage	Consensus (bp)
DENV-1	7,106,631	19,185	0.27	492.47	5,391	37,476	0.53	402.78	10,700
DENV-2	7,106,631	51,054	0.72	678.51	10,377	51,773	0.73	660.30	10,691
DENV-3	7,106,631	27,424	0.39	353.47	10,599	19,344	0.27	241.62	10,673
DENV-4	7,106,631	1,106,640	15.57	14,159.70	10,663	1,106,937	15.58	14,160.50	10,637

Abbreviations: bp, base pair.

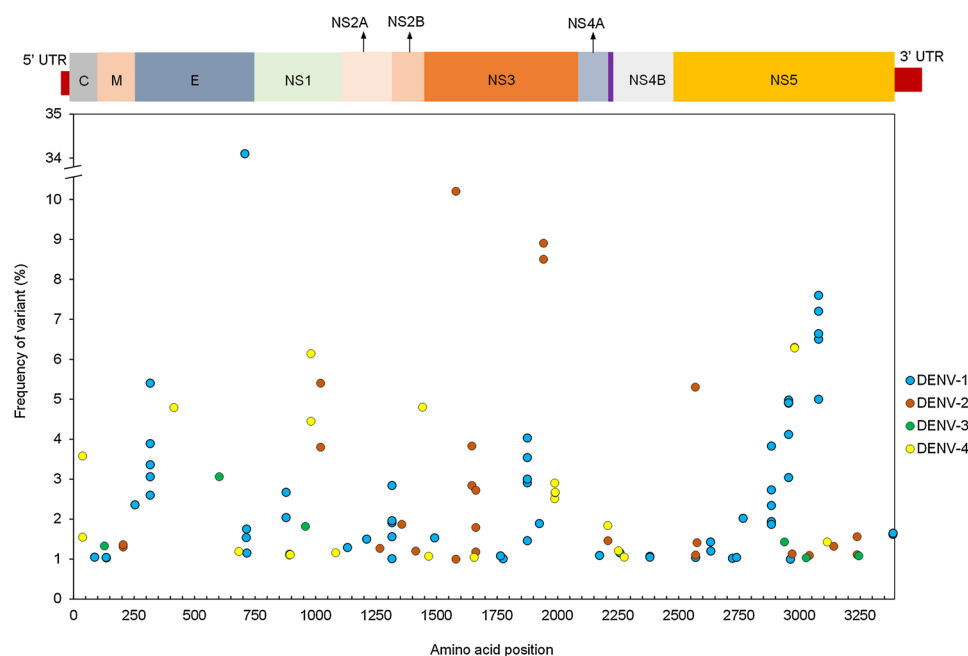


Fig. 3. Distribution of single nucleotide variants (SNVs) across the sequenced DENV genomes. Schematic representation of the DENV encoded proteins is shown above the graph. The graph depicts the frequency of SNVs by amino acid position are shown for DENV-1, DENV-2, DENV-3 and DENV-4. Every dot represents a non-synonymous change on the sequence, and the different colors in dots indicates different serotypes. Variant calling was performed in CLC Genomics Workbench v10.1.1 using the following parameters: 500-fold coverage, and a 1% base frequency (InDels and structural Variants, Q-score threshold = 30, [p-value < 0.0001]).

DNA/RNA leading to cross contamination, something that has been described previously [47]. To minimize the chance of contamination we i) used unique dual-index combinations to diminish the possibility of

misassignment on multiplexing that could generate conflicts in downstream analysis [48], ii) performed a size selection after library pooling in order to eliminate fragments below 150bp ensuring that free indexes

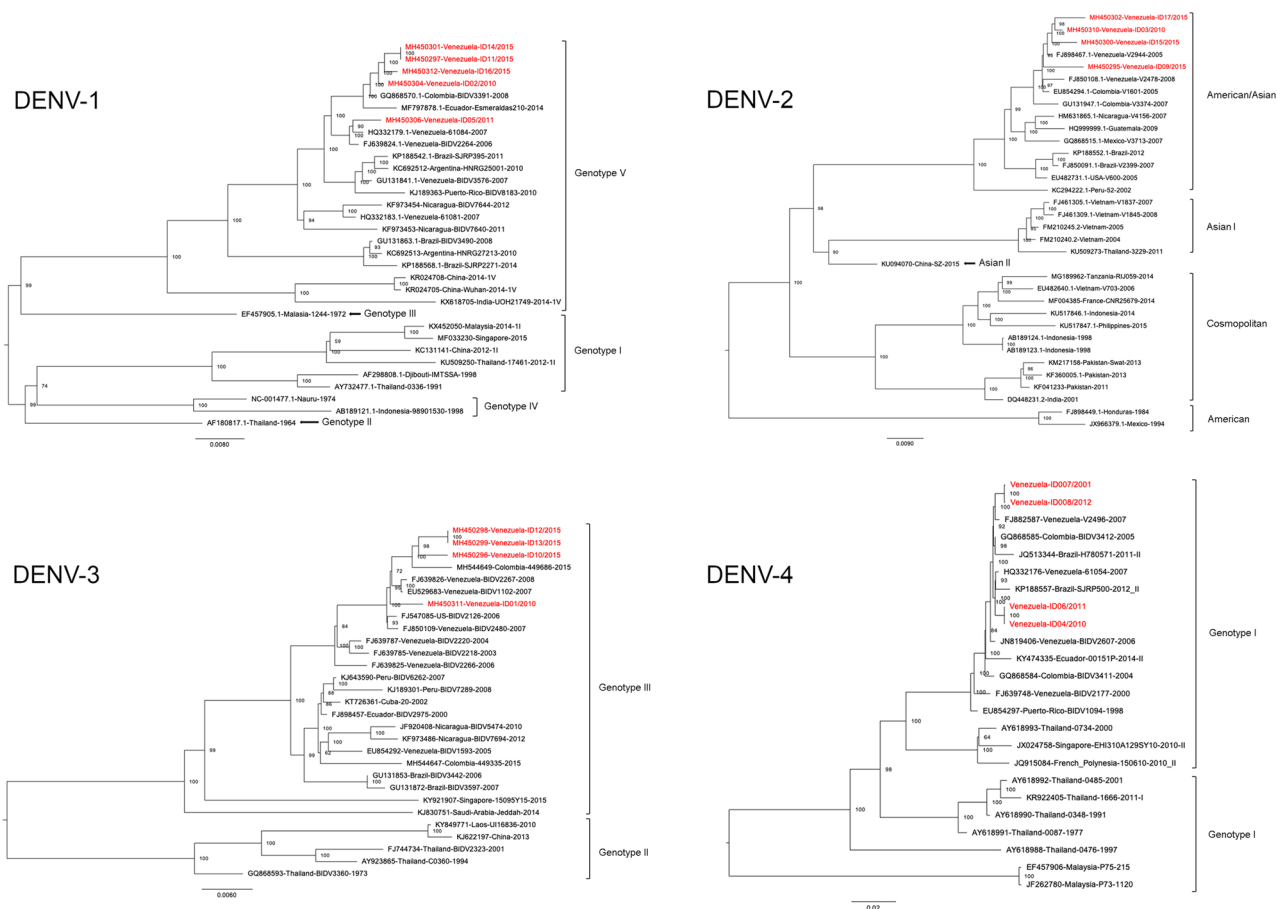


Fig. 4. Maximum Likelihood (ML) phylogenetic trees derived from the ORFs of DENV serotypes. The trees are mid-point rooted for visualization purposes. Red tips represent the sequences reported in this study. The scale bar represents the number of nucleotide substitution per site. Trees were constructed under the GTR + Γ substitution model using RAxML software with bootstrap support of 1000 replicates (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

were not present in the final libraries which also minimizes index hopping [48], iii) measured the amount of library loaded onto the flow cell to assure optimal cluster density thereby decreasing the possibility of mismatches on cluster assignment (cross talk/sample bleeding) [49] and iv) used a setting of zero barcode mismatches when using the bcl2fastq2 software to guarantee that only barcodes with 100% identity will be used during demultiplexing.

The ability to detect multiple DENV in a single sample without targeting any specific serotype, the high concordance with RT-PCR or RT-qPCR and furthermore, the possibility of multiplexing up to 24 different samples with TS and 384 different samples with NXT makes shotgun metagenomics ideal for genetic surveillance of DENV and other arboviruses without the need for a complex inventory of primers and probes for different viruses and strains. This may improve virus identification in public-health settings that need to screen multiple RNA viruses [33]. Additionally, recent Ebola and Zika striking epidemics revealed the relevance of continuous surveillance, rapid diagnosis and real-time tracking of emerging infectious diseases for containment efforts during nascent outbreaks [50], for which shotgun metagenomics may help to detect unnoticed pathogens' circulation by existing surveillance systems, e.g. Zika circulation since 2013 [51–53]. Moreover, detailed studies of complete genomes could help in the design of tailor-made assays for detection and typing of specific strains (i.e. virulent or outbreak strains) and likewise may be used to evaluate the effect of antivirals and vaccines on DENV populations, and to monitor the emergence of resistant or immune escaped mutants [54].

Conclusion

A shotgun metagenomics approach can be applied to successfully sequence whole genomes of DENV directly from clinical samples, without the need for prior sequence-specific amplification steps. This is essential for the rapid surveillance of DENV, namely to understand major epidemics and swiftly develop containment control strategies. The ability to detect infection with multiple DENV serotypes together with the high throughput of the NGS platforms could facilitate an in-depth analysis of co-viral infections and the linkage to clinical manifestations and possible association with specific strains. This could shed light into the reported relationship of inter- and intra-host DENV diversity (*quasispecies*) and human hosts. Finally, this approach can also be used for the design of vaccines against DENV in different epidemiological settings by predicting antigenic regions that are common to the circulating DENV serotypes and likewise to monitor the emergence of resistant DENV strains during vaccination campaigns.

Ethics statement

This study followed international standards for the ethical conduct of research involving human subjects. Data and sample collection were carried out within the DENVEN and IDAMS (International Research Consortium on Dengue Risk Assessment, Management and Surveillance) projects. The study was approved by the Ethics Review Committee of the Biomedical Research Institute, Carabobo University (Aval Bioetico #CBIIB(UC)-014 and CBIIB(UC)-2013-1), Maracay, Venezuela; the Ethics, Bioethics and Biodiversity Committee (CEBioBio) of the National Foundation for Science, Technology and Innovation (FONACIT) of the Ministry of Science, Technology and Innovation, Caracas, Venezuela; the regional Health authorities of Aragua state (CORPOSALUD Aragua) and Carabobo State (INSALUD); and by the Ethics Committee of the Medical Faculty of Heidelberg University and the Oxford University Tropical Research Ethics Committee.

Availability of supporting data

The data sets supporting the results of this article are included in the supplementary material data. The Illumina short read sequences are

available in SRA under the accession number [SRP149651](#) and assembled genomes are deposited in the NCBI database under accession number [MH450295-MH450312](#).

Funding

This study was supported by the Venezuelan Nacional Science, Technology and Innovation Funds (FONACIT) during data and sample collection in Venezuela [Grant Number 2011000303]; the INTERREG VA funded project EurHealth-IHealth, part of a Dutch-German cross-border network supported by the European Commission, the Dutch Ministry of Health, Welfare and Sport (VWS), the Ministry of Economy, Innovation, Digitalisation and Energy of the German Federal State of North Rhine-Westphalia and the German Federal State of Lower Saxony [Grant Number 202085]; the International Research Consortium on Dengue Risk Assessment, Management and Surveillance (IDAMS), funded by FP7-HEALTH-2011 [Grant Agreement Number 281803]. Erley Lizarazo received the Abel Tasman Talent Program grant from the UMCG, University of Groningen, Groningen, the Netherlands. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

John W. Rossen consults for IDbyDNA. All other authors declare no conflicts of interest. IDbyDNA did not have any influence on interpretation of reviewed data and conclusions drawn, nor on drafting of the manuscript and no support was obtained from them.

Acknowledgments

We thank Alberto Aguilar Briceño and Izabela Rodenhuis-Zybert for their valuable support in providing the DENV-2 strain 16,681 controls. We thank Maria Guadalupe Guzman for her valuable support in the performance of molecular detection of part of the samples.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.btecx.2019.100009>.

References

- Simmons, C.P., Farrar, J.J., Nguyen, V., Wills, B., 2012. Dengue. *N. Engl. J. Med.* 366 (15), 1423–1432. <https://doi.org/10.1056/NEJMr1110265>.
- Weaver, S.C., Vasilakis, N., 2009. Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infect. Genet. Evol.* 9 (4), 523–540. <https://doi.org/10.1016/j.meegid.2009.02.003>.
- European Centre for Disease Prevention and Control. Local Transmission of Dengue Fever in France and Spain – 2018 — 22 October 2018. ECDC, Stockholm Accessed 20 Nov 2018. <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-local-transmission-dengue-fever-france-and-spain>.
- Bhatt, S., Gething, P., Brady, O., Messina, J., Farlow, A., Moyes, C., et al., 2013. The global distribution and burden of dengue. *Nature*. 496 (7446), 504–507. <https://doi.org/10.1038/nature12060>.
- Leitmeyer, K.C., Vaughn, D.W., Watts, D.M., Salas, R., Villalobos, I., de Chacon, et al., 1999. Dengue virus structural differences that correlate with pathogenesis. *J. Virol.* 73, 4738–4747.
- Guzman, M.G., Halstead, S.B., Artsob, H., Buchy, P., Farrar, J., Gubler, D.J., et al., 2010. Dengue: a continuing global threat. *Nature Rev Microbiol.* 8, S7–S10. <https://doi.org/10.1038/nrmicro2460>.
- Azhar, E.I., Hashem, A.M., El-Kafrawy, S.A., Abol-Ela, S., Abd-Alla, A., Sohrab, S., et al., 2015. Complete genome sequencing and phylogenetic analysis of dengue type 1 virus isolated from Jeddah, Saudi Arabia. *Virol. J.* 12 (1). <https://doi.org/10.1186/s12985-014-0235-7>.
- Vaughn, D.W., Green, S., Kalayanarooj, S., Innis, B.L., Nimmannitya, S., Suntayakorn, S., et al., 2000. Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. *J. Infect. Dis.* 181 (1), 2–9. <https://doi.org/10.1086/315215>.
- Lanciotti, R.S., Calisher, C.H., Gubler, D.J., Chang, G.-J., Vorndam, A.V., 1992. Rapid detection and typing of Dengue viruses from clinical samples by using Reverse

- Transcriptase-Polymerase Chain Reaction. *J. Clin. Microbiol.* 30, 545–551.
- Wang, E., Ni, H., Xu, R., Barrett, A.D.T., Watowich, S.J., Gubler, D.J., Weaver, S.C., 2000. Evolutionary relationships of endemic/epidemic and sylvatic dengue viruses. *J. Virol.* 74, 3227–3234.
- Avilés, G., Rowe, J., Meissner, J., Manzur Caffarena, J.C., Enria, D., Jeor, S., 2002. Phylogenetic relationships of dengue-1 viruses from Argentina and Paraguay. *Arch. Virol.* 147, 2075–2087.
- Kukreti, H., Chaudhary, A., Rautela, R.S., Anand, R., Mittal, V., Chhabra, M., et al., 2008. Emergence of an independent lineage of dengue virus type 1 (DENV-1) and its co-circulation with predominant DENV-3 during the 2006 dengue fever outbreak in Delhi. *Int. J. Infect. Dis.* 12 (5), 542–549. <https://doi.org/10.1016/j.ijid.2008.02.009>.
- Rico-Hesse, R., 1990. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virol.* 174, 479–493.
- Ramos-Castañeda, J., Barreto dos Santos, F., Martinez-Vega, R., Galvão de Araujo, J.M., Joint, G., Sarti, E., 2017. Dengue in Latin America: systematic review of molecular epidemiological trends. *PLoS Neg. Trop. Dis.* 11 (1). <https://doi.org/10.1371/journal.pntd.0005224>. e0005224.
- Henn, M., Bosch, I., Harris, E. Broad Institute Microbial Genome Sequencing & Analysis. Comparative Genomics of Dengue Virus: genome population structure, transmission, and understanding differential inflammatory disease responses. Cambridge, MA 02141 USA. 2005.
- Christenbury, J.G., Aw, P.P.K., Ong, S.H., Schreiber, M.J., Chow, A., Gubler, D.J., et al., 2010. A method for full genome sequencing of all four serotypes of the dengue virus. *J. Virol. Methods* 169, 202–206. <https://doi.org/10.1016/j.jviromet.2010.06.013>.
- Baronti, C., Piorkowski, G., Leparc-Goffart, I., de Lamballerie, X., Dubot-Pérès, A., 2015. Rapid next-generation sequencing of dengue, EV-A71 and RSV-A viruses. *J. Virol. Methods* 226, 7–14. <https://doi.org/10.1016/j.jviromet.2015.09.004>.
- Rodriguez-Roche, R., Blanc, H., Borderia, A.V., Díaz, G., Henningson, R., Gonzalez, D., et al., 2016. Increasing clinical severity during a Dengue virus type 3 Cuban epidemic: deep sequencing of evolving viral populations. *J. Virol.* 90 (9), 4320–4333. <https://doi.org/10.1128/JVI.02647-15>.
- Cruz, C.D., Torre, A., Troncos, G., Lambrechts, L., Leguia, M., 2016. Targeted full-genome amplification and sequencing of dengue virus types 1–4 from South America. *J. Virol. Methods* 235, 158–167.
- Farci, P., Strazzera, R., Alter, H.J., Farci, S., Degioannis, D., Coiana, A., Peddis, G., Usai, F., Serra, G., Chessa, L., et al., 2002. Early changes in hepatitis C viral quasiespecies during interferon therapy predict the therapeutic outcome. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3081–3086. <https://doi.org/10.1016/j.jviromet.2016.06.001>.
- Lee, H.Y., Perelson, A.S., Park, S.C., Leitner, T., 2008. Dynamic correlation between in-host HIV-1 quasiespecies evolution and disease progression. *PLoS Comput. Biol.* 4, e1000240. <https://doi.org/10.1371/journal.pcbi.1000240>.
- Lauring, A.S., Andino, R., 2010. Quasiespecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6 <https://doi.org/10.1371/journal.ppat.1001005>. e1001005.
- Tyler, A.D., Christianson, S., Knox, N.C., Mabon, P., Wolfe, J., Van Domselaar, G., et al., 2016. Comparison of sample preparation methods used for the next-generation sequencing of *Mycobacterium tuberculosis*. *PLoS One* 11 <https://doi.org/10.1371/journal.pone.0148676>. e0148676.
- Lan, J.H., Yin, Y., Reed, E.F., Moua, K., Thomas, K., Zhang, Q., 2015. Impact of three Illumina library construction methods on GC bias and HLA genotype calling. *Hum. Immunol.* 76, 166–175. <https://doi.org/10.1016/j.humimm.2014.12.016>.
- Schirmer, M., D'Amore, R., Ijaz, U.Z., Hall, N., Quince, C., 2016. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinf.* 2016 (17), 125. <https://doi.org/10.1186/s12859-016-0976-y>.
- Hasan, M.R., Rawat, A., Tang, P., Jithesh, P.V., Thomas, E., Tan, R., Tilley, P., 2016. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J. Clin. Microbiol.* 54, 919–927. <https://doi.org/10.1128/JCM.03050-15>.
- Holden, K.L., Harris, E., 2004. Enhancement of dengue virus translation: role of the 3' untranslated region and the terminal 3' stem-loop domain. *Virol.* 329, 119–133. <https://doi.org/10.1016/j.virol.2004.08.004>.
- Flygare, S., Simmon, K., Miller, C., Qiao, Y., Kennedy, B., Di Sera, T., Graf, E.H., Tardif, K.D., Kapusta, A., Rynearson, S., Stockmann, C., Queen, K., Tong, S., Voelkerding, K.V., Blaschke, A., Byington, C.L., Jain, S., Pavia, A., Ampofo, K., Eilbeck, K., Marth, G., Yandell, M., Schlager, R., 2016. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* 17, 111. <https://doi.org/10.1186/s13059-016-0969-1>.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. <https://doi.org/10.1093/molbev/mst010>.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Illumina Technical Note – Optimizing Cluster Density on Illumina Sequencing Systems. <https://support.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf> (accessed 25 March, 2019).
- Uzcategui, N.Y., Camacho, D., Comach, G., Cuello de Uzcategui, R., Holmes, E.C., Gould, E.A., 2001. Molecular epidemiology of dengue type 2 virus in Venezuela: evidence for *in situ* virus evolution and recombination. *J. Gen. Virol.* 82, 2945–2953. <https://doi.org/10.1099/0022-1317-82-12-2945>.
- Svraka, S., Rosario, K., Duizer, E., van der Avoort, H., Breitbart, M., Koopmans, M., 2010. Metagenomic sequencing for virus identification in a public-health setting. *J. Gen. Virol.* 91 (11), 2846–2856. <https://doi.org/10.1099/vir.0.024612-0>.
- Nasheri, N., Petronella, N., Ronholm, J., Bidawid, S., Corneau, N., 2017. Characterization of the genomic diversity of norovirus in linked patients using a metagenomic deep sequencing approach. *Front. Microbiol.* 8, 1–14. <https://doi.org/10.3389/fmicb.2017.00073>.
- Tan, L.V., Tuyen, N.T.K., Thanh, T.T., Ngan, T.T., Van, H.M.T., Sabanathan, S., et al., 2015. A generic assay for whole-genome amplification and deep sequencing of enterovirus A71. *J. Virol. Methods* 215–6, 30–36. <https://doi.org/10.1016/j.jviromet.2015.02.011>.
- Schlager, R., Chiu, Y.C., Miller, S., Procop, G.W., Weinstock, G., 2017. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch. Pathol. Lab. Med.* 141, 776–786. <https://doi.org/10.5858/arpa.2016-0539-RA>.
- Montmayer, A.M., Ng, T.F.F., Schmidt, A., Zhao, K., Magaña, L., Iber, J., et al., 2017. High-throughput next-generation sequencing of polioviruses. *J. Clin. Microbiol.* 55, 606–615. <https://doi.org/10.1128/JCM.02121-16>.
- Houlihan, C.F., Frampton, D., Bridget Ferns, R., Raffle, J., Grant, P., Reidy, M., et al., 2018. Use of whole-genome sequencing in the investigation of a nosocomial influenza virus outbreak. *J. Infect. Dis.* 218 (9), 1485–1489. <https://doi.org/10.1093/infdis/jiy333>.
- Rodriguez-Roche, R., Villegas, E., Cook, S., Paulie, A.W., Poh, K., Hinojosa, Y., et al., 2012. Population structure of the dengue viruses, Aragua, Venezuela, 2006–2007. Insights into dengue evolution under hyperendemic transmission. *Infect. Genet. Evol.* 12 (2), 332–344. <https://doi.org/10.1016/j.meegid.2011.12.005>.
- Ramírez, A., Fajardo, A., Moros, Z., Gerder, M., Caraballo, G., Camacho, D., et al., 2010. Evolution of dengue virus type 3 genotype III in Venezuela: diversification, rates and population dynamics. *Virol. J.* 7, 329. <https://doi.org/10.1186/1743-422X-7-329>.
- King, R., 2010. The atlas of human migration. Global Patterns of People on the Move. Earthscan, United Kingdom.
- Pfeiffer, J.K., Kirkegaard, K., 2005. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog.* 1 (2), 0102–0110. <https://doi.org/10.1371/journal.ppat.0010011>.
- Choudhury, M.A., Lott, W.B., Banu, S., Cheng, A.Y., Teo, Y.Y., Ong, R.T.H., Aaskov, J., 2015. Nature and extent of genetic diversity of dengue viruses determined by 454 pyrosequencing. *PLoS One* 10 (11), 1–15. <https://doi.org/10.1371/journal.pone.0142473>.
- Aaskov, J., 2006. Long-Term Transmission of Defective RNA Viruses in Humans and Aedes Mosquitoes. *Science* 311 (5758), 236–238. <https://doi.org/10.1126/science.1115030>.
- Li, D., Lott, W.B., Lowry, K., Jones, A., Thu, H.M., Aaskov, J., 2011. Defective interfering viral particles in acute dengue infections. *PLoS One* 6 (4). <https://doi.org/10.1371/journal.pone.0019447>. e19447.
- Graf, E.H., Simmon, K.E., Tardif, K.D., Hymas, W., Flygare, S., Eilbeck, K., et al., 2016. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. *J. Clin. Microbiol.* 54, 1000–1007. <https://doi.org/10.1128/JCM.03060-15>.
- Street, T.L., Sanderson, N.D., Atkins, B.L., Brent, A.J., Cole, K., Foster, D., et al., 2017. Molecular diagnosis of orthopedic-device-related infection directly from sonication fluid by metagenomic sequencing. *J. Clin. Microbiol.* 55, 2334–2347. <https://doi.org/10.1128/JCM.00462-17>.
- Illumina Technical Note - Effects of Index Misassignment on Multiplexing and Downstream Analysis. <https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862> (accessed 25 March, 2019).
- Mitra, A., Skrzypczak, M., Ginalski, K., Rowicka, M., 2015. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform. *PLoS One* 10 (4). <https://doi.org/10.1371/journal.pone.0120520>. e0120520. Published 2015 Apr 10.
- Gardy, J.L., Loman, N.J., 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Statist. Annot.* 19, 9–20. <https://doi.org/10.1038/nrg.2017.88>.
- Faria, N.R., Azevedo, R., Kraemer, M., Souza, R., Cunha, M., Hill, S., et al., 2016. Zika virus in the Americas: Early epidemiological and genetic findings. *Science.* 352, 345–349. <https://doi.org/10.1126/science.aaf5036>.
- Faria, N.R., Quick, I., Claro, J., Theze, J., de Jesus, M., Giovanetti, M., et al., 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature.* 546, 406–410. <https://doi.org/10.1038/nature22401>.
- Grubaugh, N.D., Ladner, J., Kraemer, M., Dudas, J., Tan, A., Gangavarapu, K., et al., 2017. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature.* 546, 401–405. <https://doi.org/10.1038/nature22400>.
- Sim, S., Hibberd, M.L., 2016. Genomic approaches for understanding dengue: insights from the virus, vector, and host. *Gen. Biol.* 17, 38. <https://doi.org/10.1186/s13059-016-0907-2>.
- Cordeiro, M.T., Braga-Neto, U., Nogueira, R.M.R., Marques, E.T., 2009. A Reliable classifier to differentiate primary and secondary acute dengue infection based on IgG ELISA. *PLoS One* 4 (4), e4945. <https://doi.org/10.1371/journal.pone.0004945>.
- World Health Organization, 2009. Dengue Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition. World Health Organization, Geneva.